Calibration of long-term global horizontal irradiation estimated by HelioClim-3 through short-term local measurement campaigns: extending of the results to European and African sites

Christophe Vernay, Sébastien Pitaval, Philippe Blanc

CALIBRATION OF LONG-TERM GLOBAL HORIZONTAL IRRADIATION ESTIMATED BY HELIOCLIM-3 THROUGH SHORT-TERM LOCAL MEASUREMENT CAMPAIGNS: EXTENDING OF THE RESULTS TO EUROPEAN AND AFRICAN SITES

Christophe Vernay, Sébastien Pitaval
SOLAÏS
400, avenue Roumanille, BP 309
06906 Sophia Antipolis cedex, France
cvernay@solais.fr; spitaval@solais.fr

Philippe Blanc
MINES ParisTech, CEP - Centre énergétique et procédés
1 rue Claude Daunesse, BP 207
06904 Sophia Antipolis cedex, France
philippe.blanc@mines-paristech.fr

ABSTRACT

This paper proposes an adaptation of a recent ground-based short-term calibration algorithm applied to long-term time-series of global horizontal irradiation (GHI) provided by HelioClim-3 (HC3), a satellite-based surface solar irradiation database; it extends the initial conclusions for the South-East of France to a larger coverage. A first analysis of the long-term ground pyranometric measurements leads to the characterization of the clearness index error variability which confirms the systematic presence of, at least, a sinusoid component which period is equal to the astronomical year. On contrary of the first results based on the South-East of France, because the phasing of this sinusoid highly varies from one site to another, an adaptation of the original calibration procedure is proposed in order to have it applicable under different latitudes. The resulting mean bias error on the monthly GHI systematically goes below 3% when considering a 12-month local measurement campaign, while the seasonal variability of the error is drastically reduced.

1. INTRODUCTION

Strong attention is paid to the prediction of the yearly yield of photovoltaic (PV) projects with important economic constraints. The reliability of this prediction depends on the accuracy of both the PV system modeling and the prediction of the irradiation impinging on the module plane.

For a given geographical location, PV modeling software such as PVsyst (available at www.pvsyst.com) is able to create hourly global tilted irradiation data with respect to PV panel orientation – and other relevant meteorological data such as air temperature, using global horizontal irradiation (GHI) as input. Transposition in the collector plane is performed using different models such as Hay or Perez models (1). Monthly GHI is therefore the minimum information regarding local irradiation which is mandatory for the yield assessment of any PV project.

Different types of databases allow assessing the GHI of a specific site: databases based on ground pyranometric stations, others based on satellite images and finally derived and system integrating databases. This paper focuses on HelioClim-3 (HC3) database which is available through SoDa web service (www.soda-is.com) and which has been constructed since 2004 and daily updated since, through the processing of Meteosat Second Generation satellite images by the Heliosat-2 method (2, 3).

Even though the mean bias error (MBE) on the estimation of the GHI is very low for all satellite databases, its variability from one site to another is not negligible when considering a same database (4). Because the MBE is a systematic error that cannot be reduced by time aggregation (*e.g.* month-to-year aggregation), it has direct impact on the so-called typical irradiation (both monthly and yearly) that is used as an input to PV modeling software. It is therefore important for the companies that work in the PV field and that are end-user of the HC3 database, to try and refine this satellite-based estimation through a local calibration using on-site pyranometric measurements. Besides, it should be noted that this is already the norm and latest practices in the area of wind energy for which a short campaign (6 months to 1 year)

usually measures the local wind distribution at the early stage of the project. It is now appearing for large PV one or several pyranometers during a short period of time (typically lower than a year).

A first study has addressed the local calibration of the daily GHI estimated by HC3 for the Provence-Alpes-Côte d'Azur (PACA) region, in the South-East of France (5, 6). It has focused on nine sites located in non-mountainous area of PACA Region for which orography does not disturb the local measurement.

This paper aims at assessing the applicability of the past results to four new sites, with ground pyranometric measurements, still in non-mountainous area, but located at different latitudes in Europe and Africa, on the coverage of HC3.

Section 2 characterizes the error made by HC3 on the daily estimation of the GHI (through the clearness index) by processing the long-term measurements of the daily GHI for the four selected sites. This characterization confirms the analytic expression of the error of the GHI estimation derived by HC3 and applicable to the long-term estimation.

Section 3 finally recommends a short-term calibration algorithm and presents the resulting performances along with a seasonal analysis.

2. CHARACTERIZING THE ERROR FOR THE NEWLY SELECTED WORKING SITES

2.1 Description of the working pyranometric ground stations

These new works have been conducted over four pyranometric ground stations located in different countries in Europe (France, Switzerland, United Kingdom) and North Africa (Algeria), belonging to either the Meteo France network or the Baseline Surface Radiation Network (BSRN) (7) as shown in Fig.1 and detailed in Table 1.

projects where local irradiation is measured every day by

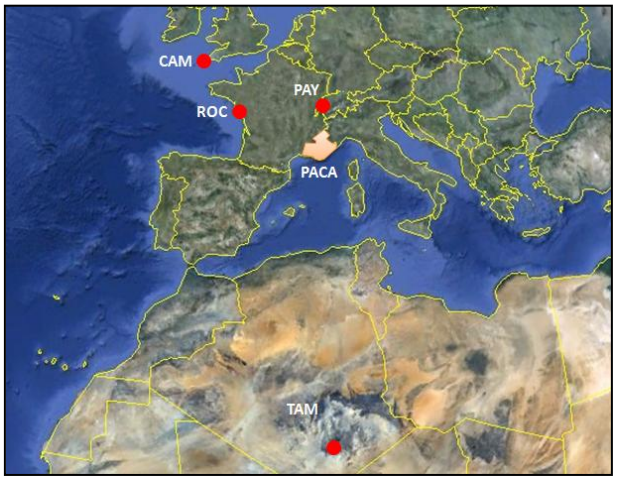


Fig. 1: Location of the four working pyranometric ground stations used for this new study, along with the location of the PACA region in the South-East of France which has been addressed in a recent study (5, 6).

TABLE 1: STATION IDENTIFICATION (ALIAS / NAME / LOCATION / ELEVATION / COUNTRY / NETWORK).

CAM	PAY	ROC	TAM
Camborne	Payerne	La Roche sur Yon	Tamanrasset
50.2167° N 5.3167° W	46.815° N 6.944° E	46.7°N 1.383 °W	22.78° N 5.51 °E
88 m	491 m	86 m	1385 m
England	Switzerland	France	Algeria
BSRN	BSRN	Météo France	BSRN

These four stations are located on non-mountainous areas (*i.e.* without noticeable shading effect from the orography) and provide directly, or thanks to aggregation procedure, the daily GHI over a long-term concomitant period with HC3 (between 3.3 and 7.5 years of daily measurements). The pyranometers are CMP6 and CMP21 from Kipp & Zonen, and PSP from Eppley. The quality of the daily GHI for the station of La Roche sur Yon belonging to the Meteo France network has been verified using the quality check protocol proposed by Geiger *et al.* (8) whereas the intra-day measurement of the GHI provided by the BSRN network (one-minute time step) has been checked following the protocol described by Roesch *et al.* (9).

2.2 Long-term performances of HC3

Table 2 depicts the performances of HC3 estimations, *i.e.* the results of the comparison between the monthly GHI estimated by HC3 and the ground measurements (the reference) for these four stations. CC is the correlation coefficient between the two series of data while the mean bias error (MBE) and the root mean square error (RMSE) used in this paper are defined as follows (considering that N is the number of monthly available data):

(Eq. 1) $MBE = \{\Sigma_i (GHI_{HC3}-GHI_{MES})\}/N$

(Eq. 2) $RMSE = \sqrt{\{\Sigma_i (GHI_{HC3}-GHI_{MES})^2/N\}}$

Both MBE and RMSE are normalized with respect to the reference value which is defined as the mean of daily measured GHI_{ST}, thus leading to nMBE and nRMSE.

TABLE 2: LONG-TERM PERFORMANCES FOR THE HC3 ESTIMATION OF MONTHLY GHI COMPARED WITH THE GROUND MEASUREMENTS FOR THE FOUR WEATHER STATIONS.

	NDATA *months*	nMBE *%*	nRMSE *%*	CC
CAM	39	1.6	8.1	0.998
PAY	71	-6.1	8.2	0.995
ROC	90	1.8	4.7	0.998
TAM	74	2.6	5.4	0.975

The correlation coefficient remains higher than 0.97 which is very satisfactory. However, the normalized MBE reaches up to -6.1% for Payerne which confirms the need to systematically refine the long-term estimation of the GHI provided by HC3 using local measurements.

2.3 Error on the clearness index

As presented in the previous study (5, 6), the daily measurements of these stations are processed in order to characterize the daily clearness index KT defined as the ratio between the daily GHI and the corresponding daily horizontal irradiation on the Top Of Atmosphere (IRR_{TOA}); KT has no unit.

(Eq. 3) $KT = GHI/IRR_{TOA}$

Because IRR_{TOA} is deterministic (it only depends on the solar constant, the distance Sun-Earth and the solar elevation angle), working on KT and working on GHI are equivalent. The rationale for working on KT relies on its systematic usage when separating the direct and diffuse components of the irradiation (10).

Fig. 2 hereafter depicts the temporal evolution of the daily error on KT between the HC3 estimation (KT_{HC3}) and the ground measurements (KT_{MES}) for the concomitant period of Camborne. The error is defined as follows:

(Eq. 4) $Error = KT_{HC3} - KT_{MES}$

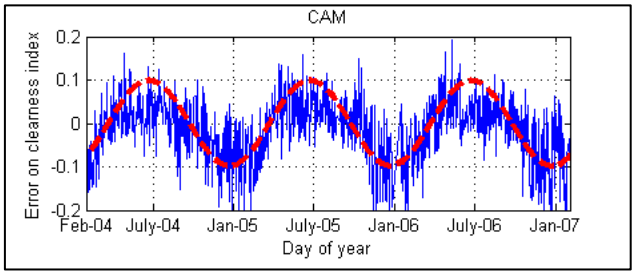
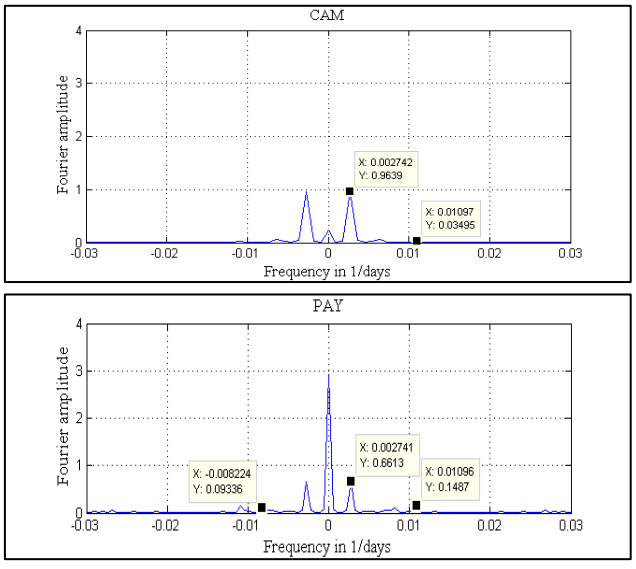


Fig. 2: Temporal analysis of the error on the clearness index for Camborne.

The temporal evolution of the error made on the clearness index highlights for the four sites the presence of a dominant sinusoidal behavior (in dotted line) even though it is much less distinguishable for Tamanrasset.

More information about the periodicity of the error can be revealed when performing a Fourier transform on a complete multi-year time series which allows decomposing it into its constituent frequencies. These spectral results are depicted on Fig. 3 where the horizontal axis indicates the value of the frequency in $days^{-1}$ and the vertical axis indicates the amplitude of the corresponding sinusoidal component (curves are symmetric with respect to the null-frequency axis, as the analyzed signals are real).



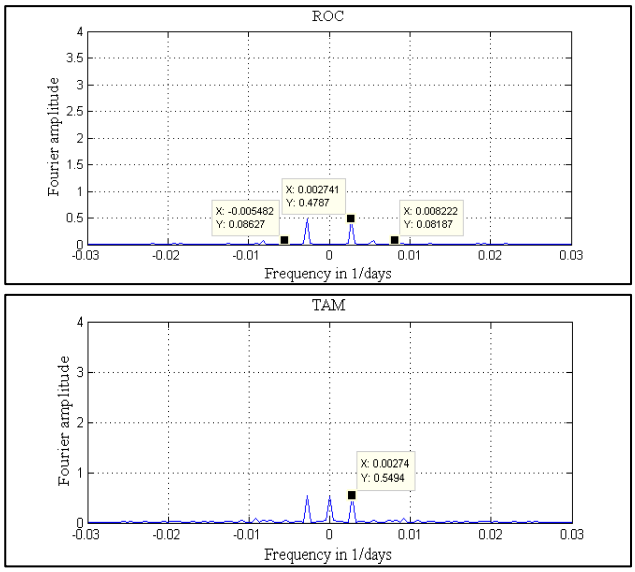


Fig. 3: The spectral analysis of the error on the clearness index confirms for each site the presence of remarkable frequencies.

As observed during the previous study performed for the PACA Region in France (5, 6), the spectral analyses confirm the presence of 2 dominant frequencies:

- F_0=0 corresponding to the bias and/or the drift of the error,
- F_1=0.00274=1/365.2422 days^{-1} which is the one-year period sinusoid depicted in dotted line on Fig.2,

plus the occasional presence of non-null frequencies:

- F_2=0.0055=2/365.2422 days^{-1},
- F_3=0.0082=3/365.2422 days^{-1},
- F_4=0.011=4/365.2422 days^{-1}.

TABLE 3: OCCURRENCE OF THE REMARKABLES FREQUENCIES WITHIN THE ERROR ON THE CLEARNESS INDEX

Weather station	Remarkable frequencies
CAM	F_0, F_1, F_4
PAY	F_0, F_1, F_3, F_4
ROC	F_1, F_2, F_3
TAM	F_0, F_1

This analysis confirms the following linear modeling of the error on the daily clearness index, applicable in the PACA Region as defined in (5, 6) as well as for the other sites studied in these new works:

(Eq. 5) $\text{Error} = KT_{HC3} - KT_{MES} = \alpha_{ST} + \beta_{ST}\, KT_{HC3} + \Sigma_i\{\gamma_{i,ST}\cos(2\pi F_i j) + \delta_{i,ST}\sin(2\pi F_i j)\}$

with:

- F_i=i/365: frequency of the sinusoid i, expressed in days^{-1}, with i ϵ {1…4};
- j: julian date defined as the decimal number of the day with the origin starting at noon Universal Time on January 1, 4713 BCE (11);
- α_{ST}, β_{ST}, $\gamma_{i,ST}$ and $\delta_{i,ST}$: coefficients that are specific to the station.

2.4 Quality of the error modeling

We have checked the quality of the modeling proposed previously at Eq. 5 for each of the four stations: a simple regression using linear least square method is performed in order to determine the coefficients which minimize the quadratic error $\|KT_{HC3}^* - KT_{MES}\|^2$ where:

(Eq. 6) $KT_{HC3}^* = \alpha_{ST} + (1-\beta_{ST})\, KT_{HC3} + \Sigma_i\{\gamma_{i,ST}\cos(2\pi F_i j) + \delta_{i,ST}\sin(2\pi F_i j)\}$

The use of both sinus and cosine for a same frequency F_i through the use of $\gamma_{i,ST}$ and $\delta_{i,ST}$ parameters allows preventing from working with the phase of the sinusoid, and therefore allows removing the non-linearity of least-square regression (5, 6).

Table 4 hereafter provides for each site the resulting nMBE and nRMSE for the monthly GHI after the calibration of the whole period of HC3 estimations (between 3.3 and 7.5 years), when considering the reduced set of frequencies {F_0 and F_1}, *i.e.* bias and/or drift plus the one-year frequency.

TABLE 4: PERFORMANCES OF THE MODELING REGARDING MONTHLY GHI COMPARED WITH GROUND MEASUREMENTS

	HC3 performance		$F_0 + F_1$	
Weather station	nMBE %	nRMSE %	nMBE %	nRMSE %
CAM	1.6	8.1	0.04	5.7
PAY	-6.1	8.2	-0.7	3.9
ROC	1.8	4.7	-0.4	3.1
TAM	2.6	5.4	0.3	2.5

These results confirm that working on the bias/drift (*i.e.* F_0) plus the F_1=1/365 days^{-1} is the cornerstone for the calibration and allows decreasing drastically the nMBE, which is very satisfactory.

However, these performances have been achieved when considering the whole period of measurement over a long-

term period, greater than 3 years. Such is not the case in the "real life" where the PV project owners cannot afford waiting for more than a year before calibrating the satellite-based estimation of the irradiation. Next section presents the calibration algorithm derived from Eq. 5 as well as the performances that can be expected for the four working sites when considering short-term measurement campaigns.

3. PERFORMANCES OF THE PROPOSED CALIBRATION CONSIDERING SHORT-TERM MEASUREMENT CAMPAIGN

3.1 Simplifying the error model determined by linear regression

The objective of such a calibration campaign is to get local measurement of irradiation on a short period (typically less than one year) using one or several pyranometers (the weather station) in order to calibrate the long-term estimation of HC3 (more than 8 years) using an *ad hoc* algorithm.

Such an algorithm can be derived from Eq. 5 which considers the 5 remarkable frequencies. Because the relative weight of frequencies F_2 to F_4 is small and their occurrence is occasional (*cf.* Fig. 3), a first approximation consists in neglecting these 3 frequencies and defining the so-called "sinus+cosine" regression which minimizes the quadratic error $||KT_{HC3}^{*}-KT_{MES}||^2$ where:

(Eq. 7) $$KT_{HC3}^{*} = \alpha_{ST} + (1-\beta_{ST})\, KT_{HC3} + \gamma_{ST} \cos(2\pi F_1 j) + \delta_{ST} \sin(2\pi F_1 j)$$

The simultaneous use of both sinus and cosine allows considering the exact phase of the 365-day sinusoid, whatever the site is, and thus allows being as close as possible to the modeling.

However, it has been shown for the PACA Region in France that the number of coefficients to determine may be reduced when an *a priori* knowledge of the phase is available (5, 6). Such is the case for all the stations located in PACA Region for which the minimum of the sinusoid is reached between December 24th and January 9th. Eq. 8 hereafter reminds the simplified regression (so-called "sinus" regression) that shall be also tested over the sites located at different latitudes:

(Eq. 8) $$KT_{HC3}^{*} = \alpha_{ST} + (1-\beta_{ST})\, KT_{HC3} + \gamma_{ST} \cos(2\pi F_1 (j-j_0))$$

with j_0 the julian date corresponding to December 31th, 2007 (the year is chosen arbitrarily).

Table 5 hereafter provides the phase of the 365-day sinusoid for each of the four sites through the occurrence date of its minimum value.

TABLE 5: OCCURRENCE DATE FOR THE MINIMUM VALUE OF THE 365-DAY SINUSOID FOR EACH OF THE FOUR SITES, ALONG WITH THE ONE FOR THE PACA REGION

Alias	Occurrence of the minimum value
CAM	December 22nd
PAY	January 27th
ROC	December 13th
TAM	November 5th
PACA (5, 6)	*Between December 24th and January 9th depending on the site. December 31th was chosen for the whole Region (cf. Eq. 8)*

The figures show that the hypothesis of the *a priori* knowledge of the phase (December 31st) is valid for the site of Camborne, rough for the sites of Payerne and La Roche sur Yon, and is finally strongly abusive for the site of Tamanrasset. It means that the phasing is not necessarily correlated with the latitude of the sites (Camborne is located at the highest latitude) which makes it difficult to determine *a priori* the phase of the dominant sinusoid (365-day period) for a site located outside PACA.

Both "sinus" and "sinus+cosine" error models have been assessed for the four sites in order to draw further conclusions.

3.2 Simulating measurement campaigns

In order to assess the performances of both error models, we have simulated for each of the four sites up to 900 measurement campaigns by extracting N consecutive daily measurements, starting from a sliding day T, among the long-term concomitant period [T0-T1] (cf. Fig. 4). This methodology allows generating a high number of measurement campaigns starting anytime in the year and whose duration varies between 1 and 12 months.

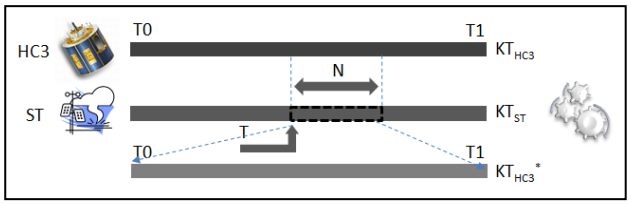


Fig. 4: Principle of the calibration of HC3 estimations through simulated measurement campaigns from ground station measurements.

Finally, for each simulated campaign, a linear regression using least square method and based on the two error models presented at Eq. 7 and Eq. 8 is performed on the short-term, in order to determine the different coefficients and assess the performance of the resulting calibration over the long-term complete time series.

3.3 Relative performances of the "sinus" and "sinus+cosine" regressions

We remind that "sinus" regression is the simplification of the "sinus+cosine" regression when forcing the phase of the 365-days sinusoid to a fix value (cf. Eq. 8). Camborne, Payerne and La Roche sur Yon sites systematically present the best performances when running the simplified "sinus" regression, whereas "sinus+cosine" regression is preferred for Tamanrasset.

This can be visualised on Fig. 5 which depicts for each site a seasonal synthesis of the relative performances of both "sinus" and "sinus+cosine" regressions applied to the simulated campaigns. It must be interpreted as follows:

- The azimuth indicates the starting month of the measurement campaign;
- The distance to the center indicates the required duration of the campaign for a given accuracy target, expressed in number of months;
- The curves indicate the minimum duration to have 95% (P95) of the simulated campaigns with |nMBE| below 5% after propagating the calibration to the long-term estimation of the GHI.

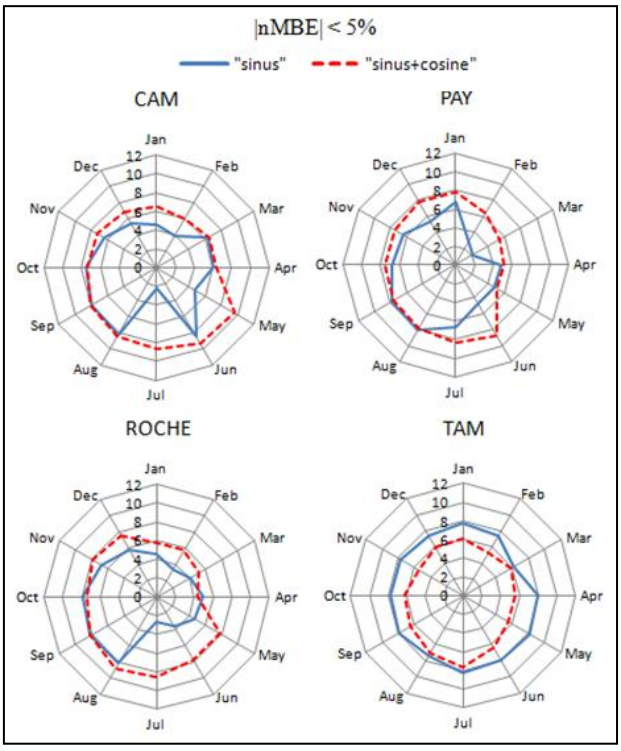


Fig. 5: P95 performances of the "sinus" (solid) and "sinus+cosine" (dotted) regressions when targeting |nMBE|<5%. The regression whose curve is located inside is better ranked as the criteria |nMBE|<5% is reached with shorter measurement campaigns.

The better performances of "sinus+cosine" regression for Tamanrasset can be explained by considering the strong gap between the phase of the 365-day sinusoid for that specific site (November 5th as shown in Table 5) and the *a priori* phase used for the "sinus" regression (December 31th). Conversely, the proximity of the site-specific phase along with the added-value of a reduced number of parameter makes the "sinus" regression more performing for the 3 other sites. Note that the preference for one regression or another is strictly the same when focusing on the 3 and 4% target: the "sinus+cosine" regression is still preferred for Tamanrasset.

Generally speaking, this leads to the conclusion that the "sinus+cosine" regression shall be preferred when *a priori* knowledge of the phase is unknown.

3.4 Performances of the "sinus+cosine" regression for the four sites.

Fig. 6 hereafter provides the performances of the "sinus+cosine" regression for the four sites along with the overall performances of the PACA Region, when considering different accuracy targets in terms of |nMBE|.

One can conclude that the individual performances of the four sites are quite similar with the overall performances of the PACA Region; this is all the more relevant when considering that more than 5300 measurement campaigns were processed for the PACA Region (9 sites) which therefore attenuates the results as only the 95% best performances are kept (P95).

These results highlight that for the four stations located at different latitudes, the "sinus+cosine" regression allows reaching |nMBE| < 3% for 95% of the measurement campaigns whose duration is 12 months. This assertion is applicable whatever the initial bias, as shown by the example of Payerne for which |nMBE| has been reduced from 6.1% down to 3%.

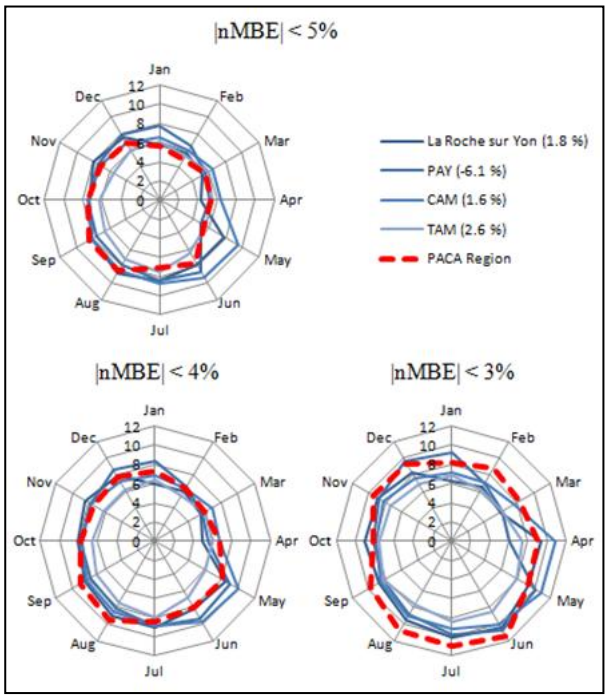


Fig. 6: P95 performances of the "sinus+cosine" regression when targeting |nMBE| lower than 5, 4 and 3%.

These latter curves present the performances of the calibration algorithms through the nMBE computed when comparing the complete data series (N months from January to December). It must be noted that such a "global" analysis of the long-term inevitably "hides" seasonal variations on the error, with inter-seasonal MBE compensation. This is illustrated by the Fig. 7 where the performances of the twelve-month campaigns are broken down by the four seasons (spring to winter). A set of three performances are depicted:

- "Initial": the HC3 raw performance where no local calibration is applied.
- "Worst case": the calibration campaign that presents the worst seasonal performances. This worst case has been determined by sorting each simulated campaigns with respect to the sum of the absolute value of nMBE by season to avoid global annual MBE compensation.
- "Average": the mean value of the seasonal nMBE when considering the whole set of twelve-month campaigns.

Each bar represents the value of nMBE in % for the corresponding period (either season or global).

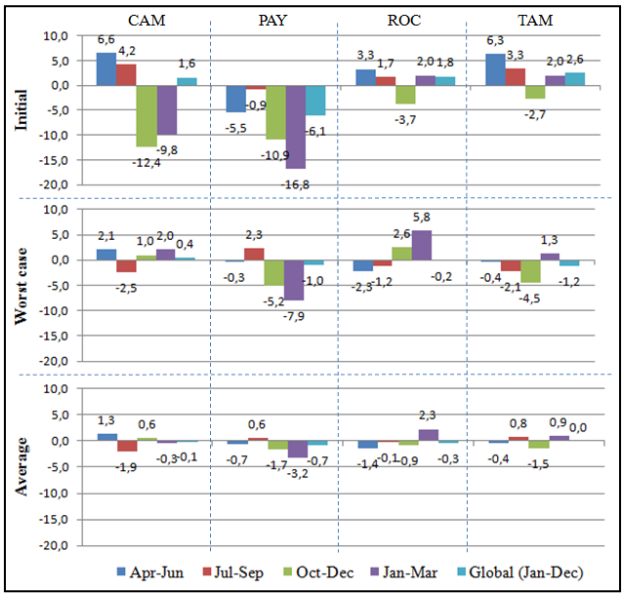


Fig. 7: Seasonal analysis of the performances of the twelve-month campaigns when using the "sinus+cosine" regression.

The "average" result shows that the "sinus+cosine" error model allows reducing drastically the nMBE for the global period as well as for each season. The "worst case" finally shows that the seasonal variability of the nMBE has been consequently reduced even though it may be degraded for some specific cases (spring/PAY, winter/ROC and autumn/TAM).

4. CONCLUSIONS

This paper has aimed at assessing the calibration approach of the GHI that was proposed by the previous study conducted over 9 stations in the same PACA Region, South-East of France (5, 6). Four new sites with pyranometric ground stations located in non-mountainous area have been selected in France, Switzerland, England and Algeria.

We have first verified that the characterization of the error made by the HelioClim-3 estimation of the clearness index is also applicable to these four new sites: each error is systematically made up of a bias and/or a drift, a dominant sinusoid of period T_1=365 days, plus other sub-frequencies $T_1/2$, $T_1/3$ and $T_1/4$ whose occurrence is site-specific.

Contrary to what was observed in PACA Region, the phase of the 365-day sinusoid may vary from one site to another, which jeopardizes a systematic simplification of the "sinus+cosine" regression to the "sinus" regression. Unless a specific local characterization is performed for a specific region (as the one made for the PACA Region), we have recommended giving priority to the "sinus+cosine" regression.

The measurements available for these four sites have allowed simulating a large number of measurement campaigns whose duration is less than 12 months and starting at different periods of the year. An analysis of the whole measurements has lead to the conclusion that the "sinus+cosine" performances for the four new sites are compliant with the performances of the sites located in the PACA Region, at different latitude. The graphical representation presented in Fig. 6 allows assessing the accuracy of this regression whatever the beginning and the duration: less than 12 month of local measurements allow reaching systematically a good accuracy ensuring |nMBE| to remain below 3% whatever the initial performance. A seasonal specific analysis of the 12-month campaigns has shown that the proposed calibration algorithm allows decreasing the seasonal variability of error on the monthly GHI.

Finally, these results presented in (5, 6) and confirmed with this new paper are the very first step regarding an innovative approach for the assessment of the local irradiation for PV projects and are run in parallel with the industrial prototyping of the *ad hoc* weather station, made of several pyranometers measuring both GHI and DHI.

ACKNOWLEDGEMENT

The contributions of data from all the various field sites to the BSRN archive is greatly appreciated.

REFERENCES

(1) Ruiz-Arias, J.A., Alsamamra, H., Tovar-Pescador, J., Pozo-Vázquez, D., 2010. Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. Energy Conversion and Management, 51, 881–893.

(2) Rigollier, C., Lefèvre, M., Wald, L., 2004. The method Heliosat-2 for deriving shortwave solar radiation from satellite images. Solar Energy, 77(2), 159-169, 2004.

(3) Blanc, Ph., Gschwind, B., Lefèvre, M., Wald, L., 2011. The HelioClim project: surface solar irradiance data for climate application. Remote Sensing, in press, 3(2), 343-361, doi:10.3390/rs20x000x

(4) Ineichen, P., 2011. Five satellite products deriving beam and global irradiance validation on data from 23 ground stations. University of Geneva, Switzerland. Available from http://www.pvsyst.com/en/publications.

(5) Vernay, C., Blanc, P., Pitaval, S., 2012. Characterizing measurement campaigns for an innovative calibration approach of the global horizontal irradiation estimated by HelioClim-3. Submitted to Solar Energy, under review. Initial manuscript available at http://www.solais.fr/Publication/CMC.pdf

(6) Vernay, C., 2011. Caractérisation des campagnes de mesures pour l'étalonnage de l'irradiation globale horizontale de la base de données satellitales HelioClim-3. Internship report. Available at http://www.solais.fr/Publication/DRSL.pdf

(7) Ohmura, A. *et al.*, 1998. Baseline Surface Radiation Network (BSRN/WCRP): New Precision Radiometry for Climate Research. Bulletin of the American Meteorological Society, vol. 79, no. 10, pp. 2115-2136, Oct. 1998.

(8) Geiger, M., Diabaté, L., Ménard, L., Wald, L., 2002. A Web service for controlling the quality of global solar irradiation. Solar Energy, 73, 6, 475-48.

(9) Roesch, A., Wild, M., Ohmura, A., Dutton, E. G., Long, C. N., Zhang, T., 2011. Assessment of BSRN radiation records for the computation of monthly means. Atmospheric Measurement Techniques, 4, 339–354.

(10) Gueymard, C.A., 2010. Progress in direct irradiance modelling and validation. Solar 2010 Conference, Phoenix, AZ, American Solar Energy Society.

(11) Doggett, L. E., 1992. Explanatory Supplement to the Astronomical Almanac. University Science Books, ed. P.K. Seidelmann.